#### Breakout Session #2 Linkage of image databases with other forms of biological knowledge

- Curtis Langlotz (Leader)
- Allan Johnson (Co-leader)
- Zhan Zhang
- John Fryman
- Yanxi Liu
- Daniel Barbara
- Kathleen Robinette
- Joseph Wang
- Fei Xu
- Ann Menkens
- Hanan Samet

- Larry Kerschberg
- Amarnath Gupta
- Beverly Meadows
- Rich Cumberlin
- Robert Tinker
- Yannick Kergosian
- John Haller
- Sanjay Ramka
- Ed Staab
- Edward Herskovits



- Definitions and context
- 3 "Use cases"
- What linkages are required
- Gap/requirements analysis
- Recommendations

# Types of Imaging Data

- Clinical images (humans)
- Small animal images
- Histology and pathology
- Anthropometrics
- 3D chemical structure
- Other microscopy (confocal, EM, etc.)
- In situ hybridization

## Non-Imaging Biological Data

- Thesauri, lexicons, vocabulary, ontology
- Medical literature, entrez databases
- Physiologic models and biochemical pathways
- Electronic medical records
- Drug databases
- Protein classification and function
- Pre-clinical
- Video surgery
- Normal human and visible human

## Use Case A: From a Clinical Radiographic Image

- What is the diagnosis given imaging features?
- What are the radiographic features to look for, given a specific diagnosis?
- What are the most salient items from the electronic medical record?
- What is optimal patient management/treatment?
- What is the molecular biology of the disease?
- What is the relevant medical literature?
- What are relevant areas for further research?

# Use Case B: From an image of a tumor model in a small animal

- What is the connection to the human model?
- What is the connection to gene expression?
- What correlative information is available from other images (histology, EM, histochemistry)?
- What is the gene expression and distribution in humans?

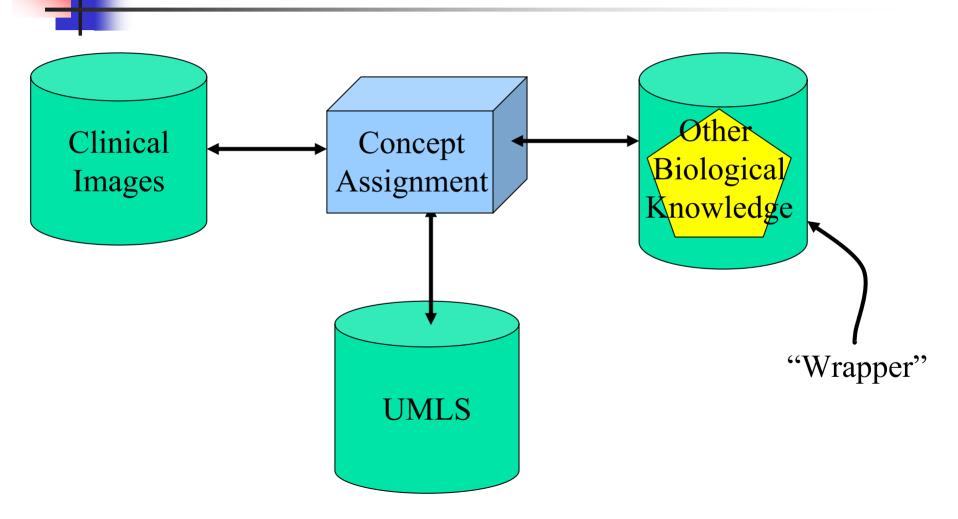
#### Use Case C: Based on a population of images of a specific disease

- More information on the disease or phenotype?
- Similar images from other databases/tumors?
- What human populations (as identified by geography, genetic markers, body shape, etc) are susceptible?

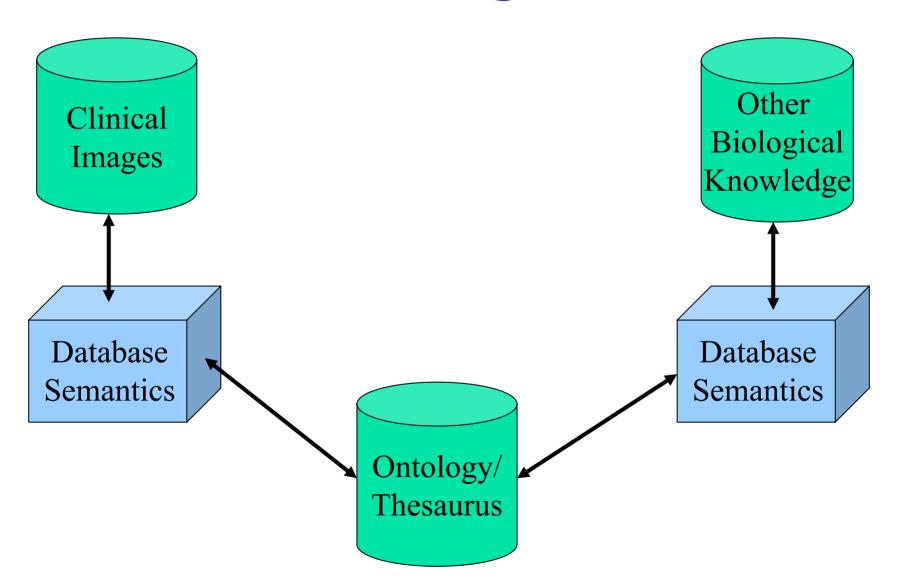
## Use Case Requirements

- Feature extraction: pixels → features, concept
  - (eg, putamen, mass effect)
- Common ontology/thesaurus of data elements, features, and concepts
- Problem solving workflow
- Automatic retrieval of similar images based on texture, shape, position.





#### Database Linkage Model





## Barriers to Linkage (1)

- No common and complete imaging data elements
  - None of the databases talk the same language: different identifiers for the same concept
  - None of the radiologists talk the same language: different meanings for the same concept
  - People and computers talk different languages: (eg, "mass effect")



#### Barriers to Linkage (2)

- Semantics (and APIs) of many existing databases are unpublished—reverse engineering required
- No thesaurus/ontology for non-clinical data (genes, proteins), internal creation and maintenance of a thesaurus required

# Recommendations

- Support efforts development and curation of imaging common data elements
- Support efforts to enhance richness of CDEs with description logic and representative images
- Support harmonized thesaurus/ontology of non-imaging data elements (UMLS for gene and protein data)
- Support development of tools to retrieve by image content